

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 08-212095

(43)Date of publication of application : 20.08.1996

(51)Int.Cl.

G06F 11/20
G06F 13/00
G06F 15/16

(21)Application number : 07-275921

(71)Applicant : HITACHI LTD

(22)Date of filing : 24.10.1995

(72)Inventor : MIYAMOTO TAKAHISA
ONO SHUJI
YASUE RIICHI
UGAJIN ATSUSHI
GOTOU NORIHIRO

(30)Priority

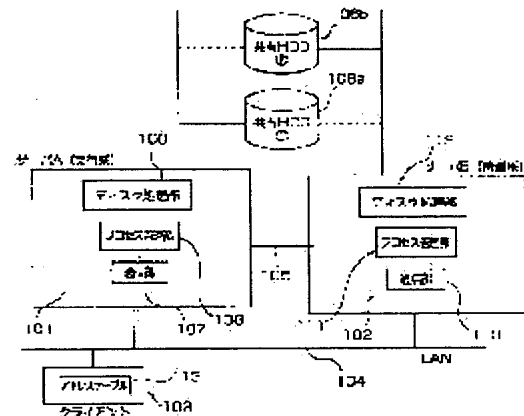
Priority number : 06267552 Priority date : 31.10.1994 Priority country : JP

(54) CLIENT SERVER CONTROL SYSTEM

(57)Abstract:

PURPOSE: To suppress the equipment/operation cost, to simplify the address management, and to shorten the suspension time of processing against occurrence of a fault in a client server control system.

CONSTITUTION: In a client server control system, a client 103 which transmits a request, a working server 101 which can have accesses to the same shared resources 106a and 106b and offers the service in response to the received request, and a stand-by server 102 which substitutes for the processing of the server 101 are connected to each other via a LAN 104. Then both servers 101 and 102 include the communication parts 107 and 110, the service processing parts 109 and 112, and the process management parts 108 and 111 respectively. When the server 101 has a fault, the request reception control of the part 107, the processing execution control of the part 109 and the request distribution control of the part 108 are totally or partly taken over by the server 102. Thus the server 102 substitutes for the service processing that should be performed by the server 101.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

rejection]

[Date of extinction of right]

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平8-212095

(43) 公開日 平成8年(1996)8月20日

(51) Int.Cl. ⁶	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 11/20	3 1 0 A			
13/00	3 0 1 P			
15/16	4 7 0 P			

審査請求 未請求 請求項の数12 O L (全 24 頁)

(21) 出願番号 特願平7-275921

(22) 出願日 平成7年(1995)10月24日

(31) 優先権主張番号 特願平6-267552

(32) 優先日 平6(1994)10月31日

(33) 優先権主張国 日本 (J P)

(71) 出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72) 発明者 宮本 貴久

神奈川県川崎市麻生区王禅寺1099番地 株

式会社日立製作所情報・通信開発本部内

(72) 発明者 大野 修司

神奈川県川崎市麻生区王禅寺1099番地 株

式会社日立製作所情報・通信開発本部内

(72) 発明者 安江 利一

神奈川県川崎市麻生区王禅寺1099番地 株

式会社日立製作所情報・通信開発本部内

(74) 代理人 弁理士 武 顕次郎

最終頁に続く

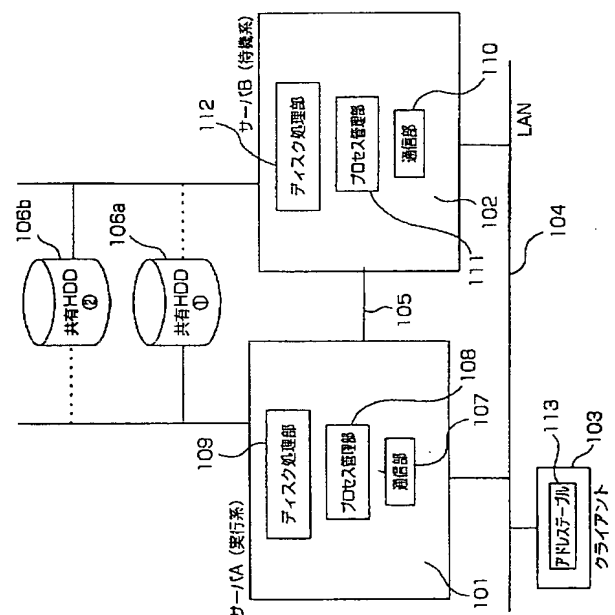
(54) 【発明の名称】 クライアントサーバ制御システム

(57) 【要約】

【課題】 クライアントサーバ制御システムに関し、従来にくらべて装備・運用コストを抑制してアドレス管理の簡便化を図り、障害発生時における処理の中断時間を短縮させる。

【解決手段】 リクエストを発信するクライアントと、同一の共有資源へのアクセスが可能で受信したリクエストに応じたサービスを提供する実行系サーバおよび実行系サーバの処理を代替する待機系サーバとが、LAN経由で相互に接続されたクライアントサーバシステムにおいて、通信部、サービス処理部、プロセス管理部を実行系サーバおよび待機系サーバにそれぞれ設ける。そして、実行系サーバ内に何らかの障害が発生した場合、通信部によるリクエストの受信制御、サービス処理部による処理の実行制御、プロセス管理部によるリクエストの配分制御の全部または一部を待機系サーバに引き継がせて、実行系サーバが行うべきサービスに関する処理を待機系サーバに代替させる。

【図 1】



1

【特許請求の範囲】

【請求項 1】 特定の処理要求が含まれたリクエストを発信する少なくともひとつのクライアントと、同一の共有資源へのアクセスが可能であって、受信したリクエスト中の処理要求に応じたサービスを発信元のクライアントに提供する実行系サーバおよび前記実行系サーバの処理を代替する待機系サーバとが、リクエストの中継伝送を行う LAN 経由で相互に接続されているクライアントサーバシステムにおいて、

クライアントから発信されたリクエストの受信制御を行う通信部と、前記共有資源へのアクセスを伴う個々のサービスに関する処理の実行制御を行うサービス処理部と、自サーバ内各部の状態を監視して障害発生時に他のサーバへ通知するとともに前記通信部で受信したリクエストの配分制御を行うプロセス管理部と、を前記実行系サーバおよび前記待機系サーバにそれぞれ設け、前記実行系サーバ内に何らかの障害が発生した場合、前記通信部によるリクエストの受信制御、前記サービス処理部による処理の実行制御、前記プロセス管理部によるリクエストの配分制御の全部または一部を前記待機系サーバに引き継がせて、前記実行系サーバが行うべき前記サービスに関する処理を前記待機系サーバに代替させることを特徴とするクライアントサーバ制御システム。

【請求項 2】 前記実行系サーバおよび前記待機系サーバのそれぞれに対して固有に割当てられている相互識別用の論理アドレスごとに前記 LAN への通信インタフェースを仮想化させる複数の仮想 LAN インタフェースと、前記仮想 LAN インタフェースを介した通信の管理制御を行う仮想 LAN インタフェース管理部と、前記実行系サーバでの障害発生あるいは正常状態への復旧に伴う前記待機系サーバとの間の各種制御の引継ぎに際して関連する論理アドレスの変更管理を行う論理アドレス管理部と、を前記実行系サーバおよび前記待機系サーバの前記通信部内に設け、通信部に障害が発生した前記実行系サーバに対応する前記待機系サーバの通信部内で、前記論理アドレス管理部は、前記実行系サーバに固有の論理アドレスに対応する予備の仮想 LAN インタフェース経由の通信を許可することを前記仮想 LAN インタフェース管理部に対して指示した後、前記 LAN に接続されているすべてのクライアントに上記許可に伴う論理アドレスの更新情報などを通知することで、前記実行系サーバが実行すべき前記サービスに関する処理を前記待機系サーバに代替させることを特徴とする請求項 1 記載のクライアントサーバ制御システム。

【請求項 3】 前記実行系サーバと前記待機系サーバとの間で行われる各種制御の引き継ぎについての管理を行う引継管理部を前記実行系サーバおよび前記待機系サーバのすべての前記サービス処理部内に設け、前記実行系サーバで障害が発生した場合、前記実行系サ

2

ーバの前記サービス処理部内に設けられた引継管理部は直前まで実行中であった前記サービスに関する処理を表す障害引継情報を前記共有資源に書き込み、前記待機系サーバの前記サービス処理部内に設けられた引継管理部は前記共有資源から読み出した前記障害引継情報に基づいて前記処理を継続実行させる代替制御を行い、前記実行系サーバが正常状態に復旧した場合、前記待機系サーバの前記サービス処理部内に設けられた引継管理部は代替実行中の前記サービスに関する処理を表す復旧引継情報を前記共有資源に書き込み、前記実行系サーバの前記サービス処理部内に設けられた引継管理部は前記共有資源から読み出した前記復旧引継情報に基づいて前記処理を継続実行させる復帰制御を行うことで、前記実行系サーバが実行すべき前記サービスに関する処理を前記待機系サーバに代替させることを特徴とする請求項 1 記載のクライアントサーバ制御システム。

【請求項 4】 常に自サーバ内各部の状態を監視している障害監視部と、前記障害監視部の監視によって得られた自サーバ内各部の状態を登録しておく状態管理テーブルと、を前記実行系サーバおよび前記待機系サーバのすべての前記プロセス管理部内に設け、前記実行系サーバで障害が発生した場合、前記障害監視部は前記実行系サーバ内の障害が発生した部分について前記プロセス管理部に通知し、これに基づいて前記プロセス管理部は前記状態管理テーブルに障害情報を登録し、既に受信されているクライアントからのリクエストを登録された前記障害情報にしたがって前記実行系サーバから前記待機系サーバに転送して、前記実行系サーバ内の障害が発生した部分に関わる処理のみを限定的に前記待機系サーバに代替させることを特徴とする請求項 1 記載のクライアントサーバ制御システム。

【請求項 5】 自サーバ内で障害が発生した部分について所定の回復処理を行う障害回復部を前記実行系サーバおよび前記待機系サーバのすべての前記プロセス管理部内に設け、前記実行系サーバで障害が発生した場合、前記障害回復部は障害が発生した部分についてのみ前記回復処理を行い、これによって前記実行系サーバが正常状態に復旧した後は再度復旧した部分が前記サービスに関する処理を行うことで、前記実行系サーバから前記待機系サーバへのリクエストの転送処理および前記待機系サーバにおける代替処理を低減させることを特徴とする請求項 4 記載のクライアントサーバ制御システム。

【請求項 6】 クライアントからのリクエストを受信順序にしたがって一時的に格納しておくリクエストキューを有するリクエスト制御部を前記実行系サーバおよび前記待機系サーバのすべての前記プロセス管理部内に設け、前記実行系サーバの前記サービス処理部で障害が発生してから前記待機系サーバの代替処理が開始されるまでの

間、

前記実行系サーバの前記プロセス管理部内に設けられたリクエスト制御部は、新たにクライアントから受信したリクエストを前記リクエストキューに格納するとともに、前記リクエストキューから順次取り出されるリクエストを前記状態管理テーブルの障害情報に基づいて前記待機系サーバに送信し、

前記待機系サーバの前記プロセス管理部内に設けられたリクエスト制御部は、クライアントまたは前記実行系サーバから受信したリクエストを前記リクエストキューに格納するとともに、前記リクエストキューから順次取り出されるリクエストを前記状態管理テーブルの障害情報に基づいて前記待機系サーバの前記サービス処理部に配分することを特徴とする請求項5記載のクライアントサーバ制御システム。

【請求項7】 前記待機系サーバに対して前記実行系サーバでの障害発生または正常状態への復旧を通知するとともにクライアントから受信したリクエストを転送するサーバ間通信部を前記実行系サーバおよび前記待機系サーバのすべての前記プロセス管理部内に設ける構成としたことを特徴とする請求項6記載のクライアントサーバ制御システム。

【請求項8】 前記実行系サーバに対して複数の前記待機系サーバを接続するとともに、前記実行系サーバの前記サービス処理部の状態と該当するサービス処理部が設けられている待機系サーバの識別子とを登録しておく状態管理テーブルを前記実行系サーバの前記プロセス管理部内に設け、

前記実行系サーバの前記サービス処理部に何らかの障害が発生した場合、前記プロセス管理部は前記状態管理テーブルの登録内容にしたがって処理を代替させる特定の待機系サーバを選択し、その待機系サーバに対して前記実行系サーバでの障害発生を通知するとともにクライアントから受信したリクエストを転送することを特徴とする請求項1記載のクライアントサーバ制御システム。

【請求項9】 前記実行系サーバおよび前記待機系サーバを直結させる専用通信線を設け、

前記実行系サーバ内に何らかの障害が発生した場合、前記プロセス管理部は前記専用通信線経由で前記待機系サーバに対して前記実行系サーバでの障害発生を通知するとともにクライアントから受信したリクエストを転送することを特徴とする請求項1あるいは4～8のいずれか一項記載のクライアントサーバ制御システム。

【請求項10】 特定の処理要求が含まれたリクエストを発信する少なくともひとつのクライアントと、同一の共有ファイル資源へのアクセスが可能であって、転送されたリクエスト中の処理要求に応じたファイルアクセスを行う少なくともひとつのファイルサーバと、クライアントから直接受信したリクエストをいずれかひとつのファイルサーバに転送する少なくともひとつのリクエスト

サーバとが、リクエストの中継伝送を行うLAN経由で相互に接続されているクライアントサーバシステムにおいて、

各々のファイルサーバごとに所定の属性状態を登録しておく状態管理テーブルを有するプロセス管理部をすべてのリクエストサーバ内に設け、

各々のリクエストサーバ内の前記プロセス管理部は、前記状態管理テーブル上に登録されている特定のファイルサーバに対してクライアントから受信したリクエストを転送し、あるファイルサーバで何らかの障害が発生した場合には、そのファイルサーバからの通知に基づき、以後のリクエストの転送先が他のファイルサーバとなるように前記状態管理テーブルの登録内容を変更することを特徴とするクライアントサーバ制御システム。

【請求項11】 あるリクエストサーバで何らかの障害が発生した場合に、そのリクエストサーバに割当てられている論理アドレスを引き継いで自己のアドレス情報およびすべてのクライアントにおけるアドレス情報を変更する通信部を設ける構成としたことを特徴とする請求項10記載のクライアントサーバ制御システム。

【請求項12】 各々のリクエストサーバ間を相互に直結させる専用通信線を設け、

あるリクエストサーバで何らかの障害が発生した場合、そのリクエストサーバ内の前記プロセス管理部は、前記専用通信線経由で他のすべてのリクエストサーバに対してそのリクエストサーバでの障害発生を通知することを特徴とする請求項11記載のクライアントサーバ制御システム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明はクライアントサーバ制御システムに係り、特に、障害発生に伴うサーバの代替制御をクライアントに全く意識させずに処理を続けるクライアントサーバ制御システムに関する。

【0002】

【従来の技術】 図22は、従来技術によるクライアントサーバ制御システムの一例の構成を概略的に示す図であり、例えばIBM社のHAFS/6000などの構成である。同図中、通常運用中のシステムでは、サーバA(201)およびサーバB(204)がともに稼働している。そして、サーバA(201)で障害が発生した場合にはサーバB(204)がサーバA(201)における処理を代替し、サーバB(204)で障害が発生した場合にはサーバA(201)がサーバB(204)における処理を代替する、ホットスタンバイ形態の運用方式を採用している。

【0003】 サーバA(201)は、クライアントA(207)からのリクエストをLAN212を介して受け付け、このリクエストに基づいて共有HDD 210に対するアクセスを行う。同様に、サーバB(204)

は、クライアントB(208)からのリクエストをLAN212を介して受け付け、このリクエストに基づいて共有HDD 211に対するアクセスを行う。ここで、一般的なLAN212では、EthernetやFDDI(Fiber Distributed Data Interface)などのインタフェースが用いられている(Ethernet:Xerox社の登録商標)。

【0004】サーバA(201)は、通常運用中にクライアントA(207)から発信されるリクエストを受信するための主アダプタA(202)と、サーバB(204)で障害が発生した場合にクライアントB(208)から発信されるリクエストをサーバB(204)の代わりに受信するための従アダプタA(203)を備えている。同様に、サーバB(204)は、通常運用中にクライアントB(208)から発信されるリクエストを受信するための主アダプタB(206)と、サーバA(201)で障害が発生した場合にクライアントA(207)から発信されるリクエストをサーバA(201)の代わりに受信するための従アダプタB(205)を備えている。さらに、サーバA(201)およびサーバB(204)は、RS-232Cなどのインタフェースを有する専用通信線209で接続されており、この専用通信線209を介して相互に障害検出や障害通知などを行っている。

【0005】図23は、図22中の各通信アダプタを識別するアドレスについての説明図である。図22のシステムで利用されている通信プロトコルのTCP/IPの場合、個々の通信アダプタに対して設定されている物理アドレスと、送信側が送信先サーバの識別に用いる論理アドレスとを組み合わせて指定することにより、単独の通信相手を決する。図23では、サーバA(201)が通常運用中に使用する主アダプタA(202)には物理アドレス“A”が、サーバA(201)には論理アドレス“1”が、それぞれ設定されている(クライアントを利用中のユーザは、論理アドレス“1”でサーバA(201)を識別する)。したがって、クライアントA(207)から物理アドレス“A”および論理アドレス“1”の組合せを指定したリクエストがLAN212に送信されると、このリクエストはサーバA(201)の主アダプタA(202)によって受信される。

【0006】図22のシステムは、サーバA(201)およびサーバB(204)が相互にバックアップし合う方式のシステムであるため、サーバA(201)の主アダプタA(202)とサーバB(204)の従アダプタB(205)に対して同一の物理アドレス“A”および論理アドレス“1”を、サーバB(204)の主アダプタB(206)とサーバA(201)の従アダプタA(203)に対して同一の物理アドレス“B”および論理アドレス“2”を、それぞれ設定しておく。

【0007】図24は、図22中のサーバA(201)に障害が発生した場合の代替制御について説明するための図である。同図中、サーバB(204)は、サーバA

(201)で障害が発生したことを専用通信線209より検知すると、従アダプタB(205)を起動させるとともに共有HDD 210の整合性チェックを行う。その後、クライアントA(207)から物理アドレス“A”および論理アドレス“1”の組合せを指定したリクエストが送信されると、代替制御によってサーバA(201)の代わりにサーバB(204)の従アダプタB(205)がこのリクエストを受信し、その内容に応じて共有HDD 210へのアクセスが行われる。代替制御中、サーバA(201)はシステム再起動などによって障害回復を図り、回復処理が完了すると専用通信線209を介して回復通知をサーバB(204)に送信する。これを受信したサーバB(204)は、従アダプタB(205)を停止させて、通常運用に復帰する。図22に示した従来のシステムでは、以上のように障害発生時の代替制御を行っている。

【0008】

【発明が解決しようとする課題】上記従来のクライアントサーバ制御システムの場合、障害発生の有無とは無関係にクライアントから次々と発信されるリクエストの受け付けを代替制御するため、複数の通信アダプタを装備するとともに障害発生時または回復時にそれらの起動または停止処理を行わなければならない、通信アダプタのアドレス管理が煩雑化するとともに装備・運用コストの抑制が困難になってしまうという問題点があった。

【0009】また、障害発生の原因がサーバを構成している各処理部の一部分であった場合でもサーバ全体の処理について別のサーバへの代替制御を行っていたことから、代替制御用のデータ群の引継ぎに要する処理時間がかなり長くなり、障害が発生したサーバで実行されていた処理の中断時間も長くなってしまいう問題点があった。特に、大容量のディスク装置が代替制御の対象になると上記による処理の中断時間は長大なものとなり、以後のクライアントサーバシステム全体における処理に多大な影響を与えてしまうという問題点があった。

【0010】したがって本発明の目的は、上記の問題点を解決して、従来にくらべて装備・運用コストを抑制してアドレス管理の簡便化を図り、障害発生時における処理の中断時間を短縮させたクライアントサーバ制御システムを提供することにある。

【0011】

【課題を解決するための手段】上記の目的を達成するために、本発明のクライアントサーバ制御システムは、特定の処理要求が含まれたリクエストを発信する少なくともひとつのクライアントと、同一の共有資源へのアクセスが可能であって、受信したリクエスト中の処理要求に応じたサービスを発信元のクライアントに提供する実行系サーバおよび前記実行系サーバの処理を代替する待機系サーバとが、リクエストの中継伝送を行うLAN経由で相互に接続されているクライアントサーバシステムに

7

において、サービス処理部、プロセス管理部、通信部を、前記実行系サーバおよび前記待機系サーバにそれぞれ設ける構成としたものである。

【0012】サービス処理部には、実行系サーバおよび待機系サーバの両方から共通にアクセス可能な共有資源として複数のインタフェースを有するHDDを接続しておくとともに、実行系サーバにおける障害発生に際して、待機系サーバによる代替処理に必要な引継情報を上記HDDに書き込み、この引継情報を読み出して処理の引き継ぎを行うための引継管理部を設ける。

【0013】通信部には、実行系サーバおよび待機系サーバのそれぞれに対して固有に割当てられている相互識別用の論理アドレスごとにLANへの通信インタフェースを仮想化させる複数の仮想LANインタフェースと、仮想LANインタフェースを介した通信の管理制御を行う仮想LANインタフェース管理部と、実行系サーバでの障害発生あるいは正常状態への復旧に伴う待機系サーバとの間の各種制御の引継ぎに際して関連する論理アドレスの変更管理を行う論理アドレス管理部と、を設ける。

【0014】プロセス管理部には、常に自サーバ内各部の状態を監視している障害監視部、障害監視部の監視によって得られた自サーバ内各部の状態を登録しておく状態管理テーブル、クライアントからのリクエストを受信順序にしたがって一時的に格納しておくリクエストキューを有してリクエストの受付およびサービス処理部への配分を行うリクエスト制御部、待機系サーバに対して実行系サーバでの障害発生または正常状態への復旧を通知する（例えば、サービス処理部で行われたHDDに対する処理についてのログデータなど（ファイル情報）や、通信部で行われた通信処理についての記録（通信情報）などを送信する）とともにクライアントから受信した未処理のリクエストを専用通信線などを介して他のサーバへ転送するサーバ間通信部、自サーバ内で障害が発生した部分について所定の回復処理を行う障害回復部などを設ける。

【0015】以下、上述した構成に基づく作用について概略的に説明する。

【0016】最初に、本発明のクライアントサーバ制御システムにおける初期化处理として、実行系サーバおよび待機系サーバにそれぞれ設けられた状態管理テーブルに、サービス処理部の状態として「正常状態」を登録しておく。

【0017】クライアントがLANを介して実行系サーバにリクエストの送信を行うと、実行系サーバの通信部がそのリクエストを受信する。そして、通信部が受信したリクエストをプロセス管理部に渡すと、プロセス管理部は状態管理テーブルの登録内容を参照して、サービス処理部の状態が「正常状態」であるときは実行系サーバにおけるサービス処理部にそのリクエストを配分し、サ

8

ービス処理部の状態が「障害発生」であるときはサーバ間通信部によりそのリクエストを専用通信線などを介して待機系サーバのプロセス管理部に配分する。

【0018】リクエストの配分を受けた実行系サーバのサービス処理部は、そのリクエストに対する処理として、共有資源であるHDD上のファイルの読み書きなどを行うとともに引継情報をHDDへ書き込む。また、プロセス管理部は実行系サーバ内の各処理部の状態を監視し、何らかの障害が発生した場合、状態管理テーブルの状態を「障害発生」に変更し、実行系サーバ上で記録されたファイル情報、通信情報、未処理のリクエストを障害発生通知として待機系サーバに送信する。

【0019】待機系サーバのプロセス管理部は、実行系サーバからファイル情報、通信情報を受信すると、これを各処理部に設定するが、実行系サーバのプロセス管理部がこれらの情報を送信できなかった場合、待機系サーバは、サービス処理部内の引継管理部により、HDDに書き込まれた引継情報からサービス処理系に関する情報を読み出す。また、実行系サーバの通信部に障害が発生している場合、待機系サーバは実行系サーバの通信部を停止させた後、論理アドレス管理部から仮想LANインタフェース管理部に対して、実行系サーバの論理アドレスを有する仮想LANインタフェースを介した通信を可能とする旨の指示を行って、待機系サーバの通信部で実行系サーバの論理アドレスが使用できるようにして、さらに、LANに接続されているすべてのクライアントに対して、上記に伴って変更されたアドレス情報を送信する。

【0020】仮想LANインタフェース管理部は、上述した論理アドレス管理部からの指示に基づいて、任意の仮想LANインタフェースを介したデータのやり取りの可／不可を管理する。なお、サーバの初期起動に際して、論理アドレス管理部は、そのサーバでデフォルトとして指定された論理アドレスについてのみ使用可とするように、仮想LANインタフェース管理部に対して指示を行う。各々のクライアントでは、受信したアドレス情報に基づいてそのクライアント上のアドレス登録情報を変更され、これによって、以後のリクエストの送信先が実行系サーバから待機系サーバに切り替わる。

【0021】実行系サーバのプロセス管理部は、障害が発生した処理部について所定の回復処理を行い、回復処理が完了すると状態管理テーブルに再度「正常状態」を登録した後、待機系サーバに対して障害回復通知を送信する。この通知を受信した待機系サーバのプロセス管理部は、それ以後実行系サーバの論理アドレスを使用しないように通信部を元に戻す。また、実行系サーバの通信部は、元に戻したアドレス情報をLANに接続されているすべてのクライアントに対して送信する。各々のクライアントでは、受信したアドレス情報に基づいてそのクライアント上のアドレス登録情報を変更され、これによ

って、以後のリクエストの送信先が待機系サーバから実行系サーバに再び切り替わる。

【0022】以上のように、本発明のクライアントサーバ制御システムは、実行系サーバに障害が発生したときに、実行系サーバの論理アドレス情報と未処理のリクエストがすべて待機系サーバに引き継がれるので、クライアントに実行系サーバの障害状況を意識させることなく、待機系サーバで実行系サーバの代替処理を行うことができる。また、実行系サーバ内に設けられている処理部ごとに状態監視、代替処理、回復処理が行われているので、障害発生時における実行系サーバから待機系サーバへの処理の引継時間、待機系サーバでの代替処理による負荷、実行系サーバでの回復処理による負荷などを最小限に抑えることができる。したがって、従来にくらべて装備・運用コストを抑制してアドレス管理の簡便化を図り、障害発生時における処理の中断時間を短縮させることができる。

【0023】

【発明の実施の形態】以下、本発明のクライアントサーバ制御システムの実施の一形態例を図面を用いて詳細に説明する。

【0024】図1は、本発明のクライアントサーバ制御システムの実施の一形態例の全体的な構成を示すブロック図である。同図中、101は実行系サーバ、102は待機系サーバ、103はクライアント、104はLAN、105はサーバ間専用通信線、106aは共有HDD、106bは共有HDD、107および110は通信部、108および111はプロセス管理部、109および112はディスク処理部である。これらのうち、共有HDD 106aと共有HDD 106bは実行系サーバ101および待機系サーバ102のいずれからもアクセス可能な請求項中の“共有資源”に、ディスク処理部109および112は請求項中の“サービス処理部”に、それぞれ相当する。なお、本明細書中に記載の待機系サーバは実行系サーバのみに専属して設けられた専用待機サーバではなく、実行系サーバが正常状態にあるときには待機系サーバ独自のサーバ処理を行っている。そして、実行系サーバに何らかの障害が発生すると、既に処理中の上記サーバ処理と並行して実行系サーバの代替処理を行う。

【0025】図1において、サーバにリクエストの送信を行うクライアント103と、クライアント103から受信したリクエストに対する処理（サービス処理）を正常状態にあるときに行う実行系サーバ101と、実行系サーバ101に障害が発生したときに実行系サーバ101の代替処理を行う待機系サーバ102とは、LAN 104を介して接続されている。実行系サーバ101と待機系サーバ102とは共有HDD 106aおよび共有HDD 106bを共有し（共有HDD 106aおよび共有HDD 106bは、アクセス用のインタフェー

スを2つ備えた「2ポートディスク」である）、そのいずれにもアクセス可能であり、サーバ相互間の通信については専用通信線105を介して行うことができる。そして、通常の運用状態では、実行系サーバ101が共有HDD 106aを、待機系サーバ102が共有HDD 106bを、それぞれ使用している。なお、以後の説明では、実行系サーバ101と待機系サーバ102とが共有および代替処理の対象とするHDDは共有HDD 106aに限られているが、共有HDD 106bについても全く問題なく共有および代替処理の対象とし得るものとする。

【0026】実行系サーバ101には、クライアント103からのリクエストを受信する通信部107と、リクエストに応じて共有HDD 106aに対するアクセスを行うディスク処理部109と、ディスク処理部109へのリクエストの配分、ディスク処理部109および通信部107における障害発生時の監視、ディスク処理部109および通信部107における障害回復処理、待機系サーバ102への障害発生通知などを行うプロセス管理部108を設ける。同様に、待機系サーバ102には、クライアント103からのリクエストを受信する通信部110と、リクエストに応じて共有HDD 106bに対するアクセスを行うディスク処理部112と、実行系サーバ101に障害が発生したときに実行系サーバ101上のプロセス管理部108から送信されたリクエストを受け付けてディスク処理部112に配分するプロセス管理部111を設ける。

【0027】図2は、図1中のプロセス管理部の内部構成を示す図である。同図中、プロセス管理部108は、リクエスト制御部501、障害監視部503、障害回復部504、サーバ間通信部505、状態管理テーブル506によって構成される。状態管理テーブル506には、クライアント103から受信したリクエストに対するサービス処理を行うディスク処理部109の状態として、サービス処理が可能なことを示す「正常状態」と、障害発生あるいはその回復処理中のためにサービス処理が不可能なことを示す「障害発生」の2種類のいずれかを登録する。リクエスト制御部501は、通信部107を介してクライアント103から受付けたリクエストを格納するリクエストキュー502を備えており、リクエストのディスク処理部109への配分、状態管理テーブル506の登録内容の更新、ファイル情報および通信情報の待機系サーバ102への通知などを行う。障害監視部503は、ディスク処理部109と通信部107における障害発生の有無を常に監視し、障害が発生したときにはリクエスト制御部501に障害発生通知を行うとともに、障害回復部504の起動処理を行う。障害回復部504は、障害が発生した処理部について所定の回復処理を行う。サーバ間通信部505は、専用通信回線105を介した実行系サーバ101から待機系サーバ102

への障害発生通知と未処理のリクエストの送信を制御する。

【0028】図3は、図2中の状態管理テーブルの具体的な構成の一例を示す図である。同図中、状態管理テーブル506は、監視処理ID部601と状態部602によって構成される。監視処理ID部601には、障害監視の対象とするディスク処理部の識別子をあらかじめ登録しておく。識別子としては、例えば、ディスク処理部固有のプロセス番号やサービス番号、ディスク処理部が配置されているメモリアドレスなどを指定する。状態部602には、障害監視の対象とするディスク処理部の現在の状態として、「正常状態」および「障害発生」のいずれかを登録する。

【0029】図4は、図1中のディスク処理部の内部構成を示す図であり、クライアント103に対して共有HDD 106aおよび共有HDD 106b（以後、両者を合わせて“共有HDD 106”と記述する）へのリモートアクセスサービスを提供しているサーバ101および102の構成要素であるディスク処理部109および112を詳細に示す。同図中、701はファイルシステム、702はブロックバッファキャッシュ、703はSCSIバスドライバ、704はSCSIバスコントローラ、705はSCSIバス、706はロギング処理部、707は主記憶、708はファイル処理引継管理部である。

【0030】図4において、ファイルシステム701は、共有HDD 106に構築されている階層構造ディレクトリ内に指定されたデータをファイル単位で格納して管理を行う。ブロックバッファキャッシュ702は、一度に読み書きされるデータ量が共有HDD 106上の格納サイズ（1セクタ＝512バイトである場合が多い）の整数倍になるように調整するとともに、他の部分にくらべて処理速度が遅いHDDへの入出力処理回数を削減するため、ファイルデータを主記憶707上に一時蓄えておくためのものである。SCSIバスドライバ703は、SCSIバス705を介して共有HDD 106とファイルデータの入出力を行うSCSIバスコントローラ704の動作制御を行う。ロギング処理部706は、ファイルシステム701によるファイル格納・管理処理の来歴情報を、共有HDD 106へのアクセスが行われる度に時々刻々共有HDD 106上へログデータ710として書き込む。ファイル処理引継管理部708（請求項中の“引継管理部”に相当する）は、サーバがクライアントに対してファイルのリモートアクセスサービスを提供するために必要な各種の設定・状態情報を共有HDD 106内に引継情報709として書き込み、実行系サーバ101で何らかの障害が発生した場合、待機系サーバ102は、この引継情報709を用いて上述したリモートアクセスサービスに関する各種の設定・状態情報を引き継ぐことにより、障害発生直前まで実行系サーバ101で処理中であつたファイル処理を継続して代替処理する。こ

で、ロギング処理部706によって共有HDD 106に書き込まれたログデータ710は、上述したファイル処理の引継ぎに際して共有HDD 106におけるディレクトリ構造およびファイル構造の整合性を簡略かつ高速に行うために利用される。

【0031】図5は、図1中の通信部の内部構成を示す図であり、クライアント103に対して共有HDD 106へのリモートアクセスサービスを提供しているサーバ101および102の構成要素である通信部107および110を詳細に示す。ここで図5の構成例は、サーバマシンとして広く利用されているUNIXワークステーションなどで採用されているTCP/IP（Transmission Control Protocol/Internet Protocol）ネットワークの一例における構成である。TCP/IPネットワークでは、ネットワーク上のノード（クライアントやサーバなど）を一意に表すためのアドレスとして、各々のLANコントローラごとに割り当てられる物理アドレスと、実際に通信処理を行うプロセス処理部が各ノード間での宛先を識別するための論理アドレスとを組み合わせる利用することによって、エンドーエンド間の通信を実現させている。したがって、同一のネットワーク内だけでなく、任意のノード経由で他のネットワーク上のノードと通信を行うことができる。

【0032】図5において、801は上述したTCP/IPネットワークのプロトコルにしたがった処理を行うプロセス処理部、802a（802b）はネットワーク内の論理アドレスごとに設けられていて上記プロセス処理部801が該当する論理アドレス806a（806b）を対象としてプロトコル処理に必要な各種の状態登録や送受信データの中継などを行うための仮想LANインタフェース、803は上記仮想LANインタフェース802を介したデータの送受信を管理する仮想LANインタフェース管理部、804は固有の物理アドレス807が割り当てられているLANコントローラ805の動作制御を行ってLAN 104経由のクライアントとの通信データの送受信を実現させるLANドライバ、808は仮想LANインタフェース管理部803に対して複数の仮想LANインタフェース802aおよび802bのいずれを通信時に利用するか指示する論理アドレス管理部である。

【0033】ここで、上述した構成の通信部を含むサーバのデフォルトアドレスとして論理アドレス 806aが設定されていた場合、サーバの立ち上げに際して論理アドレス管理部808は仮想LANインタフェース管理部803に指示を行って、LANドライバ804経由でLAN 104から受信した仮想LANインタフェース802bへの受信データと、仮想LANインタフェース802b経由で発信されるLAN 104への送信データとを通過させないようにする。一方、実行系サーバに何らかの障害が発生したために待機系サーバに論理アドレスを引き継がせた場合、論理アドレス管理部808は

仮想LANインタフェース管理部803に指示を行い、上述した仮想LANインタフェース802b経由の送受信データを通過させるようにする。

【0034】図6は、図1中のアドレステーブルの具体的な構成の一例を示す図であり、各々のクライアント上の論理アドレス806a(806b)と物理アドレス807との組合せがあらかじめ設定登録されて、TCP/IPネットワークでLAN上にデータを送信する際に利用される。すなわち、クライアント103の通信部は通信相手のサーバを論理アドレスで識別し、データ送信に際して論理アドレスのみを指定して送信先のサーバを特定するが、実際にLAN上104に送出されるデータには、図6のアドレステーブル113にあらかじめ登録されている論理アドレス806に対応した物理アドレス807が付加されているので、LAN上でのデータのやり取りは物理アドレス807によって識別することが可能である。なお、このアドレステーブルは、クライアント103だけでなく、実行系サーバ101および待機系サーバ102にも設定登録されている。

【0035】次に、図1のシステム中の実行系サーバで全機能が動作不良状態あるいは動作不可状態となってしまう完全障害が発生したときの待機系サーバへの引き継ぎ処理と、実行系サーバの再立ち上げによるシステムの正常状態への復旧処理とについて、図1を併用しながら主に図7および図8を用いて説明する。ここで、待機系サーバは、実行系サーバで障害が発生してから一定時間内に実行系サーバのプロセス管理部からファイル情報や通信情報などが通知されなかった場合に、実行系サーバで完全障害が発生したものと判断する。以下、UNIXワークステーションの標準リモートファイルアクセス機能であるNFS(Network File System)が採用されているシステムの例を対象として説明を行う。

【0036】図7は、図1のシステム中の実行系サーバで完全障害が発生したときの待機系サーバへの引き継ぎ処理の流れを示す図である。初期状態において、実行系サーバA(101)には物理アドレス“1”と論理アドレス“A”とが、待機系サーバB(102)には物理アドレス“2”と論理アドレス“B”とが、それぞれクライアント103上のアドレステーブル113に設定登録されている。クライアント103は、このアドレステーブル113の設定内容にしたがって、NFSファイルアクセスのためのリクエストや先のリクエストに対するレスポンスなどを実行系サーバA(101)との間でやり取りする。そして、やり取りを行っている間、実行系サーバA(101)と待機系サーバB(102)は、専用通信線105を介して相互に動作状態の監視を行っている。

【0037】この後、何らかの要因で実行系サーバA(101)に障害が発生した場合、待機系サーバB(102)は、上述した相互監視により実行系サーバA(1

01)の障害を検出すると、専用通信線105を介して実行系サーバA(101)に対する動作停止命令を発行する。このときの動作停止命令は、実行系サーバA(101)の状態に関わらず無条件に実行されるように、実行系サーバA(101)を構成しているハードウェアの制御信号線を直接ドライブして発行する。この動作停止命令に応じて、実行系サーバA(101)ではディスク処理部109および通信部107が閉鎖されるので、共有HDD106aやLAN104に対する一切のアクセスが停止される。しかしながら、クライアント103は、実行系サーバA(101)における障害の発生を検知する手段を持たないため、依然として実行系サーバA(101)に対してリクエストを送信するが、これに対するレスポンスが実行系サーバA(101)から得られないので、さらに同一のリクエストの再送を試みている。その間、待機系サーバB(102)のディスク処理部112は、実行系サーバA(101)が障害直前までアクセスしていた共有HDD106aにおけるディレクトリ/ファイル構造を、そのHDD内に蓄積されているログデータに基づいて調べて、動作停止命令の実行によって破壊されていた場合にはその修復を行う。

【0038】共有HDD106aにおけるディレクトリ/ファイル構造の修復が完了した後、待機系サーバB(102)のディスク処理部112におけるファイル処理引継管理部708は、共有HDD106aに書き込まれているNFS処理引き継ぎに必要な引継情報709(例えば、リモートアクセスの対象として許可された公開ディレクトリのリスト情報、障害発生時にリモートアクセスしていたクライアントに関する情報、ファイルアクセスの排他制御に関する情報など)を読み込み、新たにNFSサーバとしてサービスを開始する際に必要な初期化処理(例えば、上記公開ディレクトリのリスト情報に基づく限定アクセス処理など)を行って、NFS関連の共有情報を引き継ぐ。次に、待機系サーバB(102)の通信部110における論理アドレス管理部808は、仮想LANインタフェース管理部803に指示を行って、論理アドレス806aが付加されている仮想LANインタフェース802a経由の送受信データを通過させるようにした後、論理アドレス“A”と物理アドレス“2”とを組み合わせる旨の論理/物理アドレス変更通知を、仮想LAN10に接続されたすべてのノード(クライアント)にLANインタフェース802a経由で送信することにより、実行系サーバA(101)から論理アドレス806aを引き継ぐ処理が終了する。

【0039】上述した論理/物理アドレス変更通知を受信したクライアント103は、アドレステーブル113中の該当するエントリの更新を行う。ここで、論理アドレス806aを引き継いだ後、ファイルアクセスの排他制御に関する情報の内容によっては、ファイル処理引継管理部708が該当する状態情報の収集プロセスを起

動させて、改めてクライアントから必要な情報の収集を行う。その後、実行系サーバA(101)における障害発生で再送を繰り返していたリクエストのレスポンス待ちがタイムアウトに達すると、クライアント103は、エントリの更新を行ったアドレステーブル113に基づいて再度リクエストの送信を行うので、本来は実行系サーバA(101)に送信されるはずのリクエストが物理アドレス“2”の待機系サーバB(102)に送信されて、待機系サーバB(102)によるサービス処理を受けてからそのレスポンスが返信されるようになる。

【0040】上述した実施の形態では、実行系サーバにおける障害発生により待機系サーバに対して実行系サーバの全機能の引き継ぎが行われても、実行系サーバにリクエストを送信してサービスを受けるクライアント上のプロセスやユーザは、障害の発生や待機系サーバへの引き継ぎについて全く意識する必要がなくなる。

【0041】図8は、障害が発生した実行系サーバの再立ち上げによる図1のシステムの復旧処理の流れを示す図である。ここで、説明の前提とする実行系サーバ

(101)および待機系サーバB(102)とクライアント103の状態は、図7で説明した引き継ぎ処理後の状態であるものとする。すなわち、待機系サーバB(102)では物理アドレスとして“2”，論理アドレスとして“B”(最初から待機系サーバBに割当てられている論理アドレス)および“A”(実行系サーバAから障害発生時に引き継いだ論理アドレス)が設定され、クライアント103上のアドレステーブル113にその状態が設定登録されている。そこで、クライアント103がNFSファイルアクセスのためのリクエストを論理アドレス“A”の実行系サーバA(101)へ送信すると、そのリクエストはアドレステーブル113の登録内容にしたがって待機系サーバB(102)に受信される。そしてまた、そのリクエストに対するレスポンスも待機系サーバB(102)からクライアント103に対して送信される。

【0042】このような状態のとき、ユーザの手動操作や何らかの自動再立ち上げ手段によって実行系サーバA(101)の再立ち上げが行われると、そのことが専用通信線105を介して待機系サーバB(102)に通知される。実行系サーバA(101)から再立ち上げの通知を受けた待機系サーバB(102)は、共有HDD106aに対する新たなファイルアクセス要求をディスク処理部112で拒絶し、ディスク処理部112内のブロックバッファキャッシュ702に蓄積されているファイルデータを共有HDD106aにすべて書き出した後、共有HDD106aへのファイルアクセス処理を停止させる。次に、待機系サーバB(102)の通信部110における論理アドレス管理部808は、仮想LANインタフェース管理部803に指示を行って、論理アドレス806aが付加されている仮想LANインタフ

ェース802a経由の送受信データを通過させないようにすることにより、実行系サーバA(101)から引き継いだ通信処理を中止した後、専用通信線105を介して実行系サーバA(101)に対する再立ち上げの確認を行う。

【0043】上述した実行系サーバA(101)の再立ち上げ中、クライアント103は、実行系サーバA(101)で復旧処理が行われていることを検知する手段を持たないため、依然として実行系サーバA(101)に対してリクエストを送信するが、これに対するレスポンスが実行系サーバA(101)から得られないので、さらに同一のリクエストの再送を試みている。その間、待機系サーバB(102)から再立ち上げの確認を受けた実行系サーバA(101)では、ディスク処理部109におけるファイル処理引継管理部708が、障害発生時の引き継ぎ処理で述べたものと同様の共有情報引き継ぎ処理を行った後、通信部107における論理アドレス管理部808が、論理アドレス“A”と物理アドレス“1”とを組み合わせる旨の論理/物理アドレス変更通知を、仮想LAN10に接続されたすべてのノード(クライアント)にLANインタフェース802a経由で送信することにより、各クライアントに対して実行系サーバA(101)へのリクエスト送信を促す。

【0044】上述した論理/物理アドレス変更通知を受信したクライアント103は、アドレステーブル113中の該当するエントリの更新を行う。その後、待機系サーバB(102)への再送を繰り返していたリクエストのレスポンス待ちがタイムアウトに達すると、クライアント103は、エントリの更新を行ったアドレステーブル113に基づいて再度リクエストの送信を行うので、そのリクエストは本来の送信先である実行系サーバA(101)に送信され、サービス処理を受けてからそのレスポンスが返信されるようになる。また、待機系サーバB(102)から実行系サーバA(101)に再立ち上げの確認を行った後は、双方のサーバにおいて相互監視が再開される。

【0045】以上のように、障害が発生した実行系サーバの再立ち上げを行うことにより、待機系サーバが実行系サーバに代替して処理中であったリクエストを再度実行系サーバに復旧させる場合でも、サービス処理を受けているクライアント上のプロセスやユーザは、リクエスト処理が戻ったことや実行系サーバの復旧などについて意識する必要はない。なお、この実施の形態では、実行系サーバおよび待機系サーバの各々に対して自他の論理アドレスをあらかじめ設定しておく場合について説明を行ったが、障害が発生した時点で実行系サーバと待機系サーバとの間で論理アドレスのやり取りや設定を行うようにしてもよい。また、この実施の形態では、実行系サーバおよび待機系サーバがそれぞれ1台ずつの1対1構成について説明を行ったが、実行系サーバn台と待機系

サーバm台からなるn対m構成のシステムを構築し、ある実行系サーバで障害が発生したときにいずれか任意の待機系サーバに処理を引き継がせるように代替制御を行ってもよい。

【0046】図9～図14は、図1中の実行系サーバにおけるディスク処理部および通信部で障害が発生したときの代替制御について説明するための図である。以下、実行系サーバ101に実装されているディスク処理部109および通信部107で障害が発生した場合について、図1中のプロセス管理部108で行われる処理を中心に図9～図14を用いて説明する。なお、ディスク処理部109における詳細な処理については図4、通信部107における詳細な処理については図5および図6で説明したものと同様とする。

【0047】最初に図9に基づき、「正常状態」の実行系サーバ101における処理について説明する。同図中、通信部107がリクエスト制御部1201に対してリクエストを渡すと、リクエスト制御部1201はこれをリクエストキュー1202に格納する。そして、状態管理テーブル1206を参照してディスク処理部109に関する状態を求める。図9で、ディスク処理部109の状態が「正常状態」なので、リクエスト制御部1201は受け取ったリクエストをそのまま実行系サーバ101上のディスク処理部109に配分する。リクエストを配分されたディスク処理部109は、共有HDD 106aに記録されているデータの読み書きなどの上記リクエストに応じた処理を行う。また、障害監視部1203は、ディスク処理部109および通信部107で障害が発生していないかどうかを常に監視する。

【0048】次に図10に基づき、実行系サーバ101中のディスク処理部109および通信部107に何らかの障害が発生したときの実行系サーバ101における処理について説明する。同図中、障害監視部1203は、監視対象のディスク処理部109および通信部107のいずれかに障害が発生すると、障害回復部1204を起動させるとともに、リクエスト制御部1201への障害通知を行う。障害通知を受信したリクエスト制御部1201は、状態管理テーブル1206の状態設定を行って、障害が発生したディスク処理部109あるいは通信部107に関する状態部の設定登録値を「正常状態」から「障害発生」に変更させる。また、リクエスト制御部1201は、サーバ間通信部1205および専用通信線105を介して、待機系サーバ102に障害発生通知を送信する。送信する内容は、実行系サーバ101における障害が発生したディスク処理部109のファイル情報（障害発生時のファイルシステムの構造やNFSマウント情報など）、通信部107で障害が発生したことを示す情報、リクエストキュー1202に格納されている未処理のリクエストなどである。

【0049】次に図11に基づき、実行系サーバ101

で障害が発生した直後の待機系サーバ102における処理について説明する。同図中、専用通信線105およびサーバ間通信部1305を介して実行系サーバ101から送信されてきた障害発生通知は、リクエスト制御部1401によって受信される。障害発生通知のうち、未処理のリクエストはリクエストキュー1402に格納され、各々の処理部に対する設定が完了した後にディスク処理部112に渡される。また、ファイル情報はリクエスト制御部1401によってディスク処理部112に設定され、これによってディスク処理部112が実行系サーバ101の代替処理を実行できるようになる。さらに、リクエスト制御部1401は、通信情報に基づいて実行系サーバ101の通信部107の論理アドレスを待機系サーバ102が自動的に使用するように通信部110におけるアドレス情報の更新を行い、通信部110はLAN上のすべてのクライアントに対して更新したアドレス情報を送信する。各々のクライアントが上記アドレス情報を受信すると、クライアント上のアドレステーブルに登録されているアドレス情報がすべて更新され、リクエストの送信先は実行系サーバ101から待機系サーバ102へ自動的に切り替わる。

【0050】次に図12に基づき、障害が発生した実行系サーバ101で回復処理を行っている間の待機系サーバ102における処理について説明する。同図中、通信部110がリクエスト制御部1401にリクエストを渡すと、リクエスト制御部1401はこれをリクエストキュー1402に格納する。そして、状態管理テーブル1406を参照してディスク処理部112に関する状態を求める。図12では、ディスク処理部112の状態が「正常状態」なので、リクエスト制御部1401は受け取ったリクエストをディスク処理部112に配分する。リクエストを配分されたディスク処理部112はそのリクエストに対する処理として、共有HDD 106aに記録されているデータの読み書きなどを行う。

【0051】次に図13に基づき、障害が発生したディスク処理部109および通信部107に対する回復処理が完了するまでの実行系サーバ101における処理について説明する。同図中、障害回復部1204は、ディスク処理部109および通信部107の障害回復処理を行う。そして、回復処理が完了すると、障害回復部1204は、リクエスト制御部1201に対して回復通知を行う。回復通知を受けたリクエスト制御部1201は、状態管理テーブル1206の状態設定を行って、回復処理対象のディスク処理部1201あるいは通信部107に関する状態を、「障害発生」から「正常状態」に設定登録する。この後、リクエスト制御部1201は、サーバ間通信部1205および専用通信線105を介して障害回復通知を待機系サーバ102に送信する。最後に、通信部107は実行系サーバ102のアドレス情報をLAN上のすべてのクライアントに送信する。その結果、各

クライアント上のアドレステーブルはすべて障害発生前の状態に戻るので、リクエストの送信先は待機系サーバ 102 から実行系サーバ 101 に自動的に切り替わる。

【0052】最後に、障害回復通知を受信した待機系サーバ 102 では、リクエスト制御部 1401 が実行系サーバ 101 中の通信部 107 に対応する論理アドレスを待機系サーバ 102 が使用しないように通信部 110 の更新処理を行う。

【0053】以上のように、実行系サーバ 101 が正常状態→障害発生→障害回復→正常状態→……を繰り返すと、障害発生から障害回復までの間、本来実行系サーバ 101 で処理すべきクライアントからのリクエストを待機系サーバ 102 が受信し、実行系サーバ 101 と同様のサービスを提供する代替制御が行われる。したがって、実行系サーバ 101 で障害が発生したことをクライアントが意識することなく、リクエストの送信相手が実行系サーバ 101 から待機系サーバ 102 に自動的に変更される。また、クライアント 103 から送信されたリクエストはリクエスト制御部 1201 内のリクエストキュー 1202 に一旦格納されるので、実行系サーバ 101 から待機系サーバ 102 に障害発生通知を行なっている間でも、受信したリクエストを不用意に廃棄してしまうことを防止できる。さらに、待機系サーバ 102 に対する障害発生通知の送信と同時に、実行系サーバ 101 における障害が発生した処理部の回復処理を行うことにより、回復処理に伴ってサーバ処理が中断してしまうことを回避し、回復処理中に待機系サーバ 102 が代替していた処理を障害回復後に再び実行系サーバ 101 で処理するようにできる。

【0054】図 15～図 17 は、図 1 中の実行系サーバにおけるディスク処理部のみで障害が発生したときの代替制御について説明するための図であり、図 15 はディスク処理部 109 が正常なときの処理を、図 16 はディスク処理部 109 で障害が発生したときの処理を、図 17 はディスク処理部 109 が障害回復中のときの処理を、それぞれ示す。なお、図中のプロセス管理部、ディスク処理部、通信部の詳細については、図 2～図 6 で説明したものと同様とする。

【0055】最初に図 15 に基づき、実行系サーバ A (101) におけるディスク処理部 109 が正常なときの処理について説明する。同図中、クライアント 103 が LAN 104 を介して実行系サーバ A (101) に対してリクエストを送信すると、実行系サーバ A (101) の通信部 107 がこれを受信する。そして、通信部 107 が受信したリクエストをプロセス管理部 108 に渡すと、プロセス管理部 108 はディスク処理部 109 にリクエストを配分する。リクエストを配分されたディスク処理部 109 は、そのリクエストに対する処理として共有 HDD 106a に対するデータの読み書きなどを行う。また、プロセス管理部 108 は、ディスク処理

部 109 と通信部 107 に障害が発生していないかどうかを常に監視する。

【0056】次に図 16 に基づき、実行系サーバ A (101) におけるディスク処理部 109 で障害が発生した直後の処理について説明する。同図中、ディスク処理部 109 で障害が発生すると、プロセス管理部 108 は専用通信線 105 を通じて、ファイル情報および未処理のリクエストを障害発生通知として待機系サーバ B (102) のプロセス管理部 111 に送信する。ここでファイル情報とは、共有 HDD 106a を待機系サーバ B (102) からアクセスする際に必要となる障害発生時のファイルシステムの構造や NFS マウント情報などである。実行系サーバ A (101) から上述した障害発生通知を受信した待機系サーバ B (102) のプロセス管理部 111 は、これに基づいてディスク処理部 112 に対して実行系サーバ A (101) のファイル情報を設定する。

【0057】最後に図 17 に基づき、実行系サーバ A (101) におけるディスク処理部 109 が障害回復中のときの処理について説明する。同図中、実行系サーバ A (101) のプロセス管理部 108 は、「障害発生」状態にあるディスク処理部 109 の障害回復処理を行う。回復処理中にディスク処理部 109 に対するリクエストが実行系サーバ A (101) に到着した場合、プロセス管理部 108 が専用通信線 105 を介して待機系サーバ B (102) のプロセス管理部 111 にそのリクエストを送信する。実行系サーバ A (101) から上記リクエストを受信した待機系サーバ B (102) のプロセス管理部 111 は、ディスク処理部 112 にそのリクエストを配分する。リクエストを配分されたディスク処理部 112 は、そのリクエストに対する処理として共有 HDD 106a に対するデータの読み書きなどを行う。実行系サーバ A (101) のプロセス管理部 108 によるディスク処理部 109 の障害回復処理がすべて完了すると実行系サーバ A (101) は図 15 の「正常状態」に戻る。プロセス管理部 108 は、回復後にクライアント 103 から受信したすべてのリクエストを、実行系サーバ A (101) のディスク処理部 109 に配分する。

【0058】以上のように、実行系サーバ A (101) が正常状態→障害発生→障害回復→正常状態→……を繰り返すと、障害発生から障害回復までの間、実行系サーバ A (101) が受信したクライアントからのリクエストは専用通信線 105 を介して待機系サーバ B (102) に送信され、待機系サーバ B (102) のディスク処理部 112 によって実行系サーバ A (101) と同様のサービスが行われる。したがって、実行系サーバ A (101) で障害が発生しても、クライアント 103 がリクエストの送信相手を実行系サーバ A (101) から待機系サーバ B (102) に変更する必要がなくなる。

また、クライアント 103 から送信されたリクエストはリクエスト制御部 1201 内のリクエストキュー 1202 に一旦格納されるので、実行系サーバ A (101) から待機系サーバ B (102) に処理情報の引継ぎなどを行なっている間でも、受信したリクエストを不用意に廃棄してしまうことを防止できる。さらに、待機系サーバ B (102) に対して受信したリクエストを送信すると同時に、実行系サーバ A (101) のディスク処理部 109 の回復処理を行うことにより、回復処理に伴ってサーバ処理が中断してしまうことを回避し、回復処理中に待機系サーバ B (102) が代替していた処理を障害回復後に再び実行系サーバ A (101) で処理するよう

【0059】図 18 は、図 1 中の実行系サーバにおけるディスク処理部および専用通信線で同時に障害が発生したときの代替制御について説明するための図である。以下、実行系サーバ A (101)、待機系サーバ B (102)、クライアント 103 のいずれも接続されている LAN 104 により、リクエストに対する処理の代替制御を行う場合について説明する。なお、図中のプロセス管理部、ディスク処理部、通信部の詳細については、図 2

【0060】図 18 において、専用通信線 105 と実行系サーバ A (101) のディスク処理部 109 で同時に障害が発生した場合、実行系サーバ A (101) のプロセス管理部 108 は通信部 107 に対して、すでにクライアント 103 から受信したリクエストを実行系サーバ A (101) から待機系サーバ B (102) に転送する (ルーティング) 設定を行う。そして、実行系サーバ A (101) の通信部 107 は、障害が発生したディスク処理部 109 に対するリクエストを新たに受信すると、これをプロセス管理部 108 に渡さずに、LAN 104 を介して待機系サーバ B (102) に送信する。このリクエストを実行系サーバ A (101) から受信した待機系サーバ B (102) の通信部 110 は、プロセス管理部 111 にリクエストを渡す。そして、プロセス管理部 111 はディスク処理部 112 にリクエストを配分し、ディスク処理部 112 がそのリクエストに応じて共有 HDD、106a の読み書きなどを行う。以上のように、LAN 104 を介して実行系サーバ A (101) から待機系サーバ B (102) へリクエストを転送することにより、専用通信線 105 に障害が発生した場合でも、サーバ処理が中断してしまうことを回避できる。

【0061】図 19 ～図 21 は、本発明のクライアントサーバ制御システムの実施の他の形態例の構成および処理について説明するための図であり、クライアントからのリクエストを受信するリクエストサーバと共有 HDD に対するアクセスを行うファイルサーバとを設けて相互にバグアップさせる場合を示す。なお、図中のプロセス管理部、ディスク処理部、通信部の詳細については、

図 2 ～図 6 で説明したものと同様とする。また、状態管理テーブル中の監視処理 ID 部にはファイルサーバの識別子 (論理アドレスやホストネームなど) を、状態部には「正常状態」、「障害発生」、「待機状態」のいずれかを設定登録するものとする。

【0062】最初に図 19 に基づき、この実施の形態のシステムの構成を説明する。同図中、クライアント A (2201) はアドレステーブル 2212 を、クライアント B (2202) はアドレステーブル 2215 を、それぞれ備えており、対応するアドレステーブルの内容にしたがって、リクエストをリクエストサーバ A (2203) またはリクエストサーバ B (2204) へ送信する。LAN (2208) は、クライアント A (2201) およびクライアント B (2202) と、リクエストサーバ A (2203) およびリクエストサーバ B (2204) の間を相互に接続させる。リクエストサーバ A (2203) は通信部 2213 およびプロセス管理部 2214 を、リクエストサーバ B (2204) は通信部 2216 およびプロセス管理部 2217 を、それぞれ備えており、クライアント A (2201) およびクライアント B (2202) から送信されてきたリクエストの送受信処理を行う。プロセス管理部 2214 および 2217 は、それぞれファイルサーバ A (2205) およびファイルサーバ B (2206) の状態を登録しておく状態管理テーブルを備えている。システムが通常稼動中である場合、リクエストサーバ A (2203) のプロセス管理部 2214 における状態管理テーブルにはファイルサーバ A (2205) が「正常状態」、ファイルサーバ B (2206) が「待機状態」にあることが設定登録され、リクエストサーバ B (2204) のプロセス管理部 2217 における状態管理テーブルにはファイルサーバ A (2205) が「待機状態」、ファイルサーバ B (2206) が「正常状態」にあることが設定登録される。LAN (2210) はファイルサーバ A (2205) およびファイルサーバ B (2206) とリクエストサーバ A (2203) とを、LAN (2209) はファイルサーバ A (2205) およびファイルサーバ B (2206) とリクエストサーバ B (2204) とを、それぞれ相互に接続させる。ファイルサーバ A (2205) およびファイルサーバ B (2206) は、リクエストサーバ A (2203) またはリクエストサーバ B (2204) から送信されたリクエストを受信し、そのリクエストに応じた処理を共有 HDD (2207a) または共有 HDD (2207b) に対して実行する。ここで、共有 HDD (2207a) および共有 HDD (2207b) は、ファイルサーバ A (2205) およびファイルサーバ B (2206) のいずれからアクセス可能である。また、リクエストサーバ A (2203) とリクエストサーバ B (2204) は専用通信線 2211 を介して接続されており、両サーバ間の通信は、この専用通信線 2211 を介して行われる。

【0063】次に通常稼動中のシステムにおける処理について、図 19 に基づいて説明を行う。クライアント A (2201) は、アドレステーブル 2212 に基づいてリクエス

トをリクエストサーバA (2203) に送信する。リクエストサーバA (2203) の通信部2213は、LAN (2208) を介して送信されたリクエストを受信して、これをプロセス管理部2214に渡す。プロセス管理部2214は、状態監視テーブルを調べて、状態部の設定が「正常状態」となっているファイルサーバA (2205) にLAN (2210) を介して上記リクエストを送信する。このリクエストを受信したファイルサーバA (2205) は、リクエストの内容に応じて共有HDD 2207aに対する処理を行う。上記と同様、クライアントB (2202) は、アドレステーブル2215に基づいてリクエストをリクエストサーバB (2204) に送信する。リクエストサーバB (2204) の通信部2216は、LAN (2208) を介して送信されたリクエストを受信して、これをプロセス管理部2217に渡す。プロセス管理部2217は、状態監視テーブルを調べて、状態部の設定が「正常状態」となっているファイルサーバB (2206) にLAN (2209) を介して上記リクエストを送信する。このリクエストを受信したファイルサーバB (2206) は、リクエストの内容に応じて共有HDD (2207b) に対する処理を行う。

【0064】次に図20に基づいて、ファイルサーバA (2205) で障害が発生した場合についての説明を行う。同図中、ファイルサーバA (2205) は障害が発生したことを検知すると、LAN (2210) を介してリクエストサーバA (2203) に障害通知を送信するとともに、障害が発生した処理部の回復処理を行う。障害通知を受信したリクエストサーバA (2203) のプロセス管理部2214は、状態管理テーブルのファイルサーバA (2205) に対応する状態部の設定登録内容を「正常状態」から「障害発生」に、ファイルサーバB (2206) に対応する状態部の設定登録内容を「待機状態」から「正常状態」に、それぞれ更新する。これによって、リクエストサーバA (2203) が受信したリクエストはLAN (2210) を介してファイルサーバB (2206) に送信されるようになり、ファイルサーバA (2205) に対する待機系サーバとして、ファイルサーバB (2206) がリクエストの内容に応じて共有HDD 2207aに対する処理を行う。また、クライアントB (2202) が送信したリクエストは、図19で説明したように、リクエストサーバB (2204) およびファイルサーバB (2206) によって処理される。

【0065】最後に図21に基づいて、リクエストサーバA (2203) で障害が発生した場合についての説明を行う。同図中、リクエストサーバA (2203) は障害が発生したことを検知すると、専用通信線2211を介してリクエストサーバB (2204) に障害通知を送信するとともに、障害が発生した処理部の回復処理を行う。障害通知を受信したリクエストサーバB (2204) の通信部2216は、リクエストサーバA (2203) の通信部2213の論理アドレス (図6参照) を引き継ぐ。そして、リクエストサーバB (2204) の通信部2216が、更新されたアドレス情報をL

AN (2208) に接続されたすべてのクライアントに送信すると、これを受信したクライアントのアドレステーブルは上記と同様に更新される。これによって、クライアントA (2201) は更新後のアドレステーブル2212の設定登録内容に基づき、本来はリクエストサーバA (2203) に送信すべきすべてのリクエストを、リクエストサーバA (2203) に対する待機系サーバであるリクエストサーバB (2204) に送信するようになる。リクエストサーバB (2204) は、このようなリクエストを受信すると、これをファイルサーバA (2205) に送信する。そして、このリクエストを受信したファイルサーバA (2205) は、リクエストの内容にしたがって共有HDD 2207aに対する処理を行う。また、クライアントB (2202) が送信したリクエストは、図19で説明したように、リクエストサーバB (2204) およびファイルサーバB (2206) によって処理される。

【0066】なお、上述した実施の形態ではファイルサーバあるいはリクエストサーバに対する待機系サーバが1台である構成について説明を行ったが、待機系サーバが2台以上となるように多数のファイルサーバおよびリクエストサーバを相互に接続して、各々のファイルサーバおよびリクエストサーバの状態管理テーブルに待機系サーバとして動作させる他のファイルサーバおよびリクエストサーバの情報 (ディスク処理部の実装状態やアドレス情報など) を加えて、ファイルサーバあるいはリクエストサーバに何らかの障害があった場合には、上述した2台以上の他のファイルサーバおよびリクエストサーバのいずれかに待機系サーバとしての代替制御を実行させることにより、ファイルサーバおよびリクエストサーバによる処理の信頼性をさらに向上させることが可能となる。また、待機系サーバとして動作させるファイルサーバあるいはリクエストサーバへのリクエストの送信については、専用通信線に代えて、ファイルサーバ、リクエストサーバ、クライアントを相互に接続させているLANを介して行うようにしてもよい。

【0067】

【発明の効果】以上詳しく説明したように、本発明のクライアントサーバ制御システムによれば、実行系サーバに障害が発生したときに、実行系サーバの論理アドレス情報と未処理のリクエストがすべて待機系サーバに引き継がれるので、クライアントに実行系サーバの障害状況を意識させることなく、待機系サーバで実行系サーバの代替処理を行うことができるという効果が得られる。また、実行系サーバ内に設けられている処理部ごとに状態監視、代替処理、回復処理が行われているので、障害発生時における実行系サーバから待機系サーバへの処理の引継時間、待機系サーバでの代替処理による負荷、実行系サーバでの回復処理による負荷などを最小限に抑えることができるという効果が得られる。したがって、従来にくらべて装備・運用コストを抑制してアドレス管理の

簡便化を図り、障害発生時における処理の中断時間を短縮させることができるという効果が得られる。

【図面の簡単な説明】

【図 1】本発明のクライアントサーバ制御システムの実施の一形態例の全体的な構成を示すブロック図である。

【図 2】図 1 中のプロセス管理部の内部構成を示す図である。

【図 3】図 2 中の状態管理テーブルの具体的な構成の一例を示す図である。

【図 4】図 1 中のディスク処理部の内部構成を示す図である。

【図 5】図 1 中の通信部の内部構成を示す図である。

【図 6】図 1 中のアドレステーブルの具体的な構成の一例を示す図である。

【図 7】図 1 のシステム中の実行系サーバで完全障害が発生したときの待機系サーバへの引き継ぎ処理の流れを示す図である。

【図 8】障害が発生した実行系サーバの再立ち上げによる図 1 のシステムの復旧処理の流れを示す図である。

【図 9】図 1 中の実行系サーバにおけるディスク処理部および通信部で障害が発生したときの代替制御について説明するための図（その 1）である。

【図 10】図 1 中の実行系サーバにおけるディスク処理部および通信部で障害が発生したときの代替制御について説明するための図（その 2）である。

【図 11】図 1 中の実行系サーバにおけるディスク処理部および通信部で障害が発生したときの代替制御について説明するための図（その 3）である。

【図 12】図 1 中の実行系サーバにおけるディスク処理部および通信部で障害が発生したときの代替制御について説明するための図（その 4）である。

【図 13】図 1 中の実行系サーバにおけるディスク処理部および通信部で障害が発生したときの代替制御について説明するための図（その 5）である。

【図 14】図 1 中の実行系サーバにおけるディスク処理部および通信部で障害が発生したときの代替制御について説明するための図（その 6）である。

【図 15】図 1 中の実行系サーバにおけるディスク処理

部のみで障害が発生したときの代替制御について説明するための図（その 1）である。

【図 16】図 1 中の実行系サーバにおけるディスク処理部のみで障害が発生したときの代替制御について説明するための図（その 2）である。

【図 17】図 1 中の実行系サーバにおけるディスク処理部のみで障害が発生したときの代替制御について説明するための図（その 3）である。

【図 18】図 1 中の実行系サーバにおけるディスク処理部および専用通信線で同時に障害が発生したときの代替制御について説明するための図である。

【図 19】本発明のクライアントサーバ制御システムの実施の他の形態例の構成および処理について説明するための図（その 1）である。

【図 20】本発明のクライアントサーバ制御システムの実施の他の形態例の構成および処理について説明するための図（その 2）である。

【図 21】本発明のクライアントサーバ制御システムの実施の他の形態例の構成および処理について説明するための図（その 3）である。

【図 22】従来技術によるクライアントサーバ制御システムの一例の構成を概略的に示す図である。

【図 23】図 22 中の各通信アダプタを識別するアドレスについての説明図である。

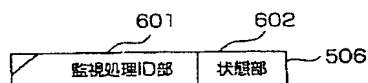
【図 24】図 22 中のサーバ A に障害が発生した場合の代替制御について説明するための図である。

【符号の説明】

- 101 実行系サーバ
- 102 待機系サーバ
- 103 クライアント
- 104 LAN（ローカルエリアネットワーク）
- 105 サーバ間専用通信線
- 106 a 共有 HDD
- 106 b 共有 HDD
- 107, 110 通信部
- 108, 111 プロセス管理部
- 109, 112 ディスク処理部

【図 3】

【図 3】



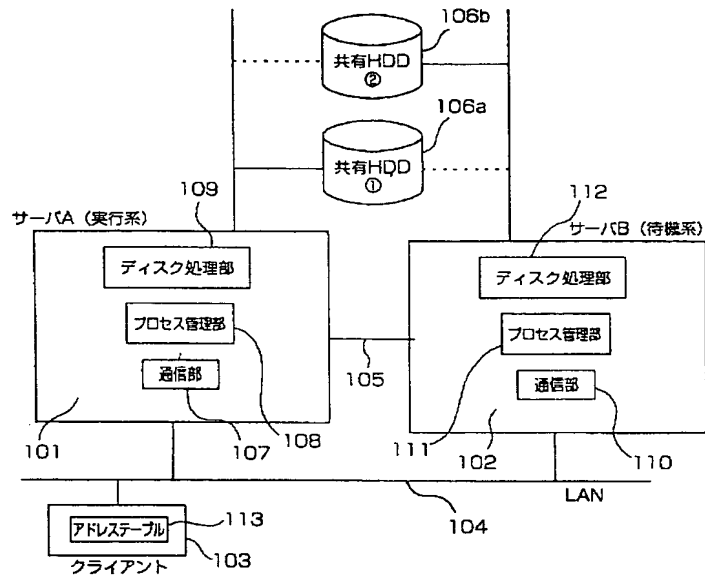
【図 6】

【図 6】

806		807
論理アドレス		物理アドレス
A		1
B		2
K		6

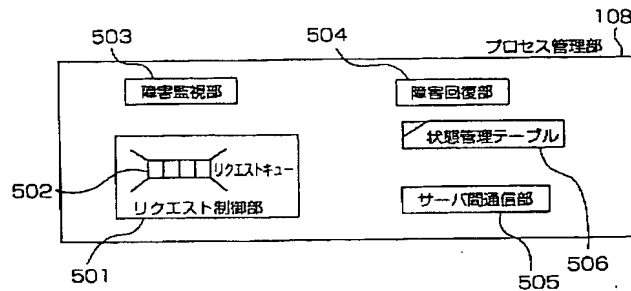
113

【図1】

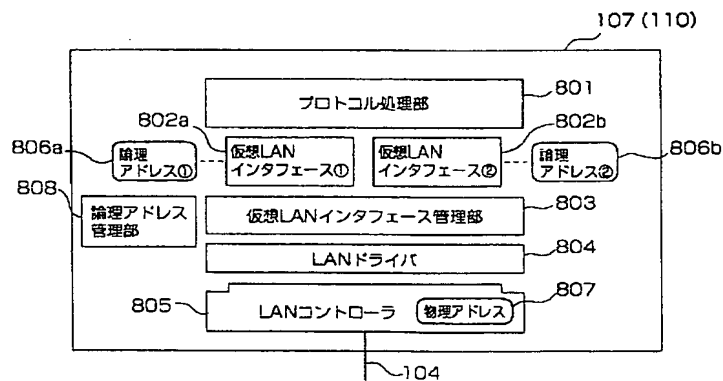


【図2】

【図2】



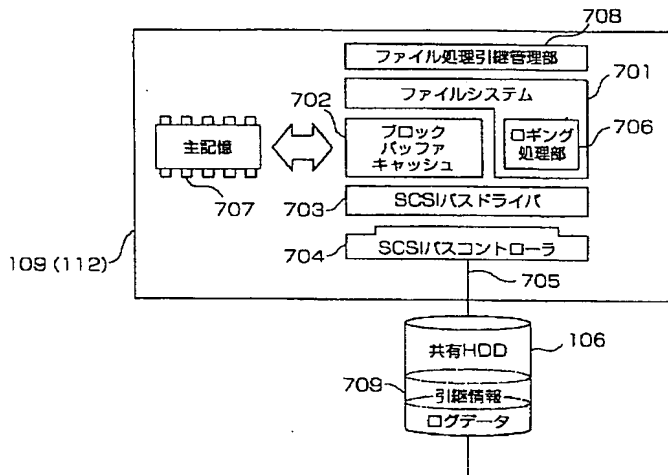
【図5】



【図1】

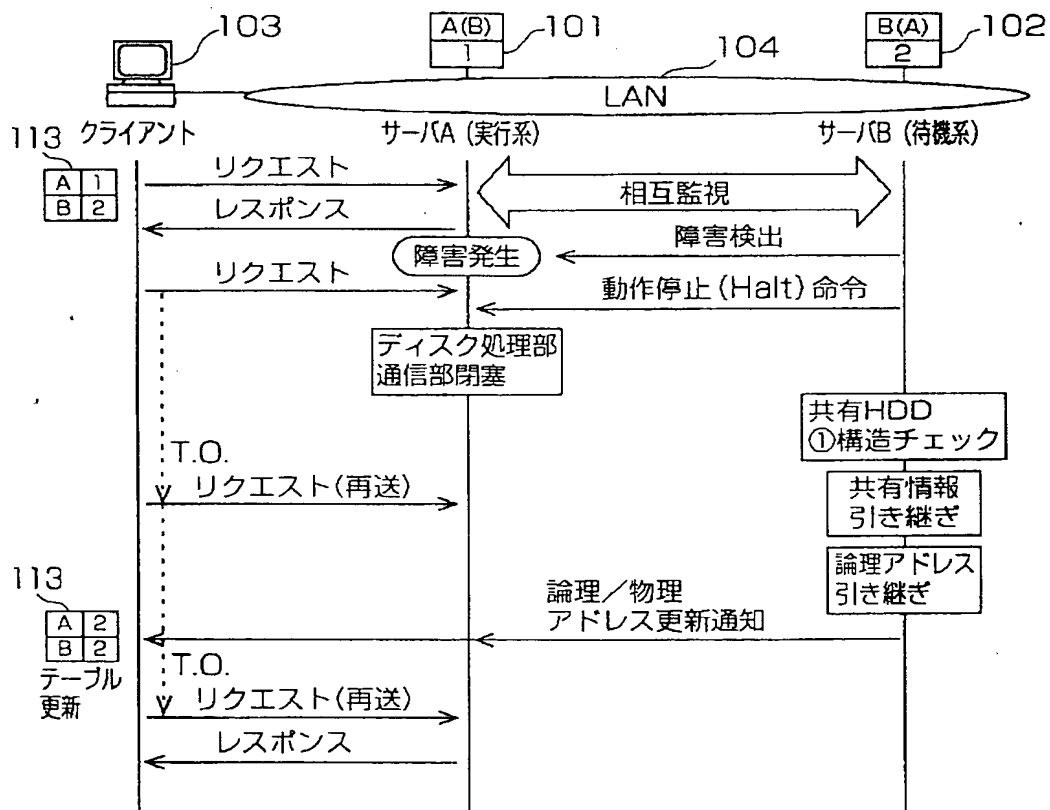
【図5】

【図 4】



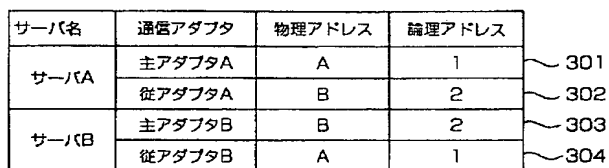
【図 7】

【図 7】

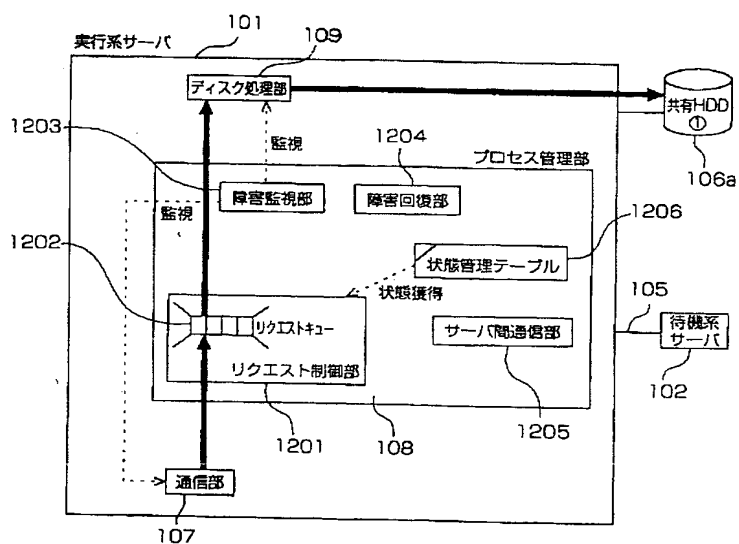


【图 23】

【圖 2 3】

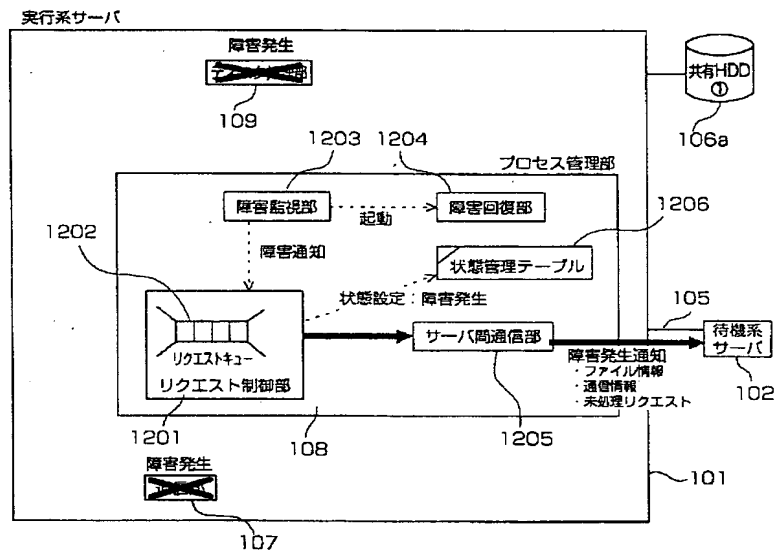


【图9】

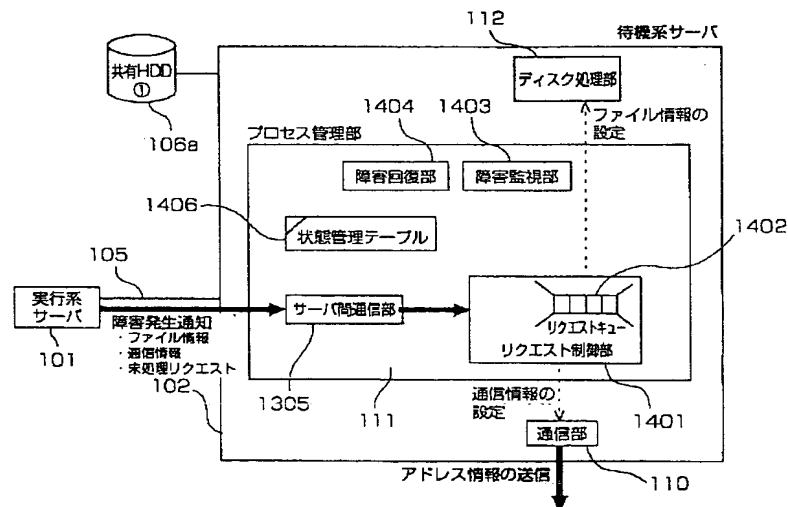


【6 図】

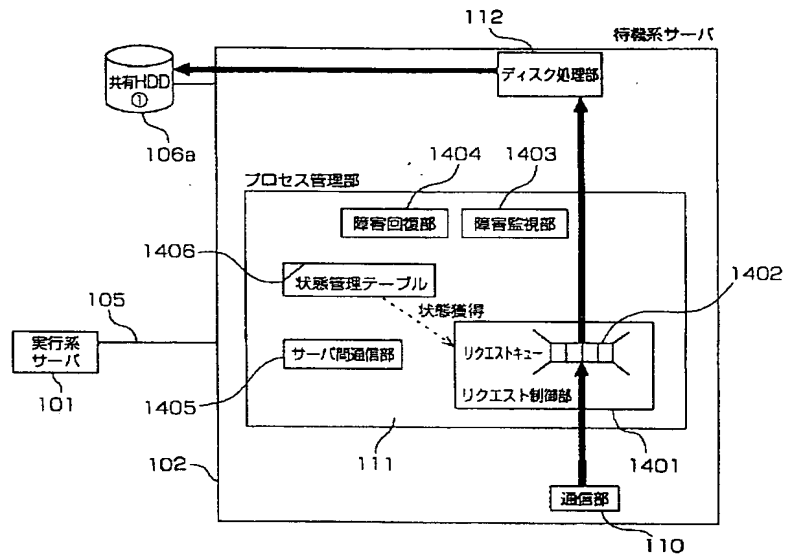
【図 10】



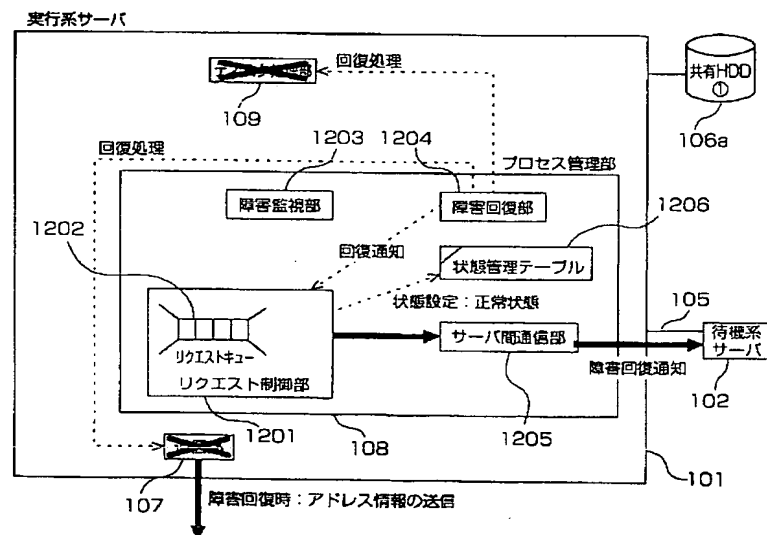
【図 11】



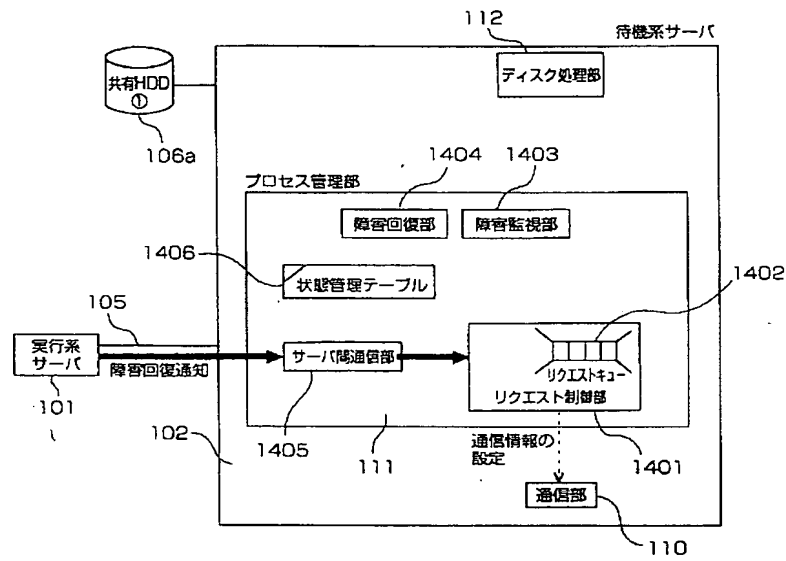
【圖 12】



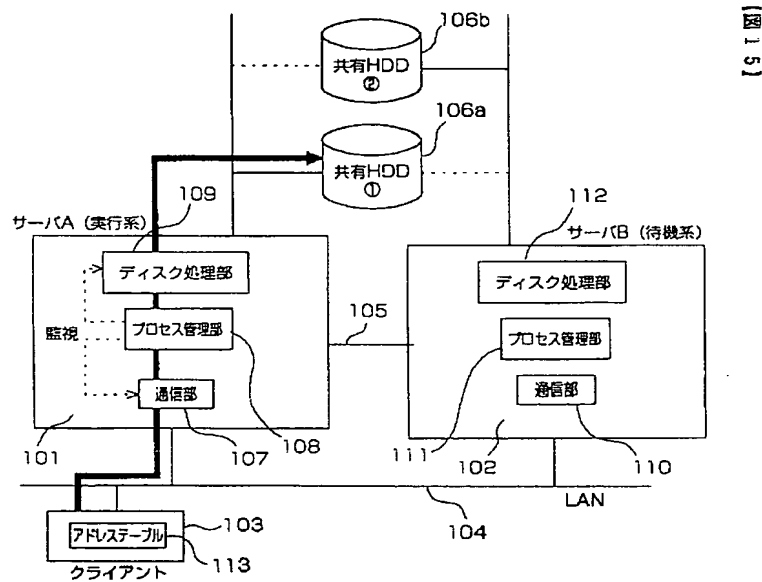
【例 13】



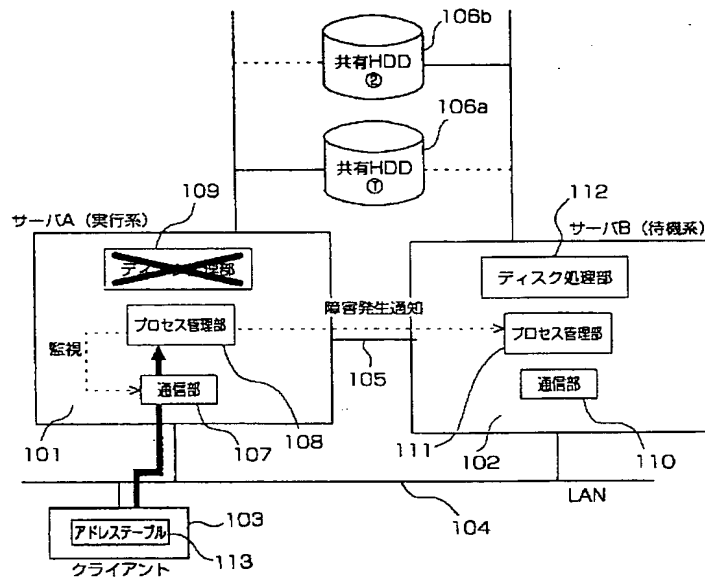
【図14】



【図15】

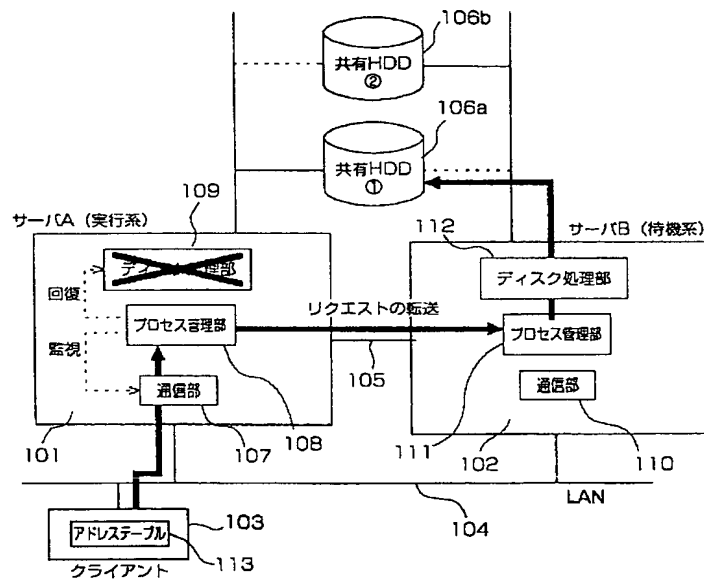


【図 16】



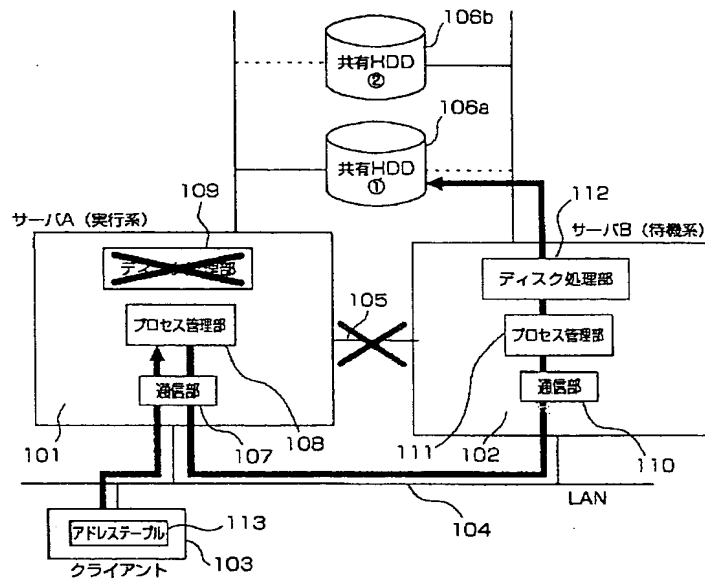
【図 16】

【図 17】



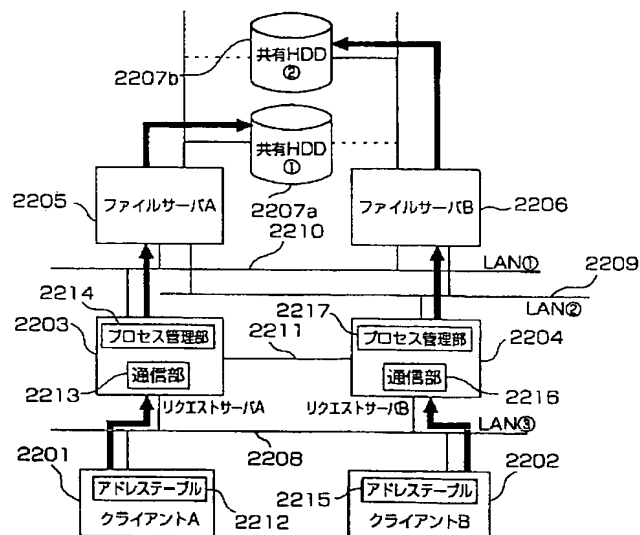
【図 17】

【図18】



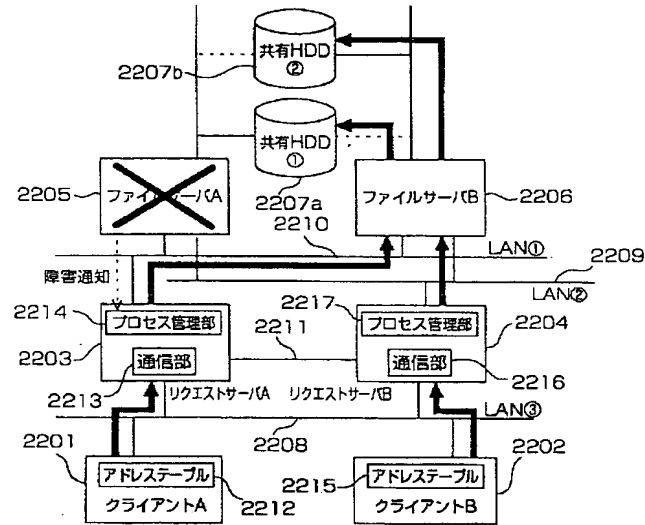
【図18】

【図19】



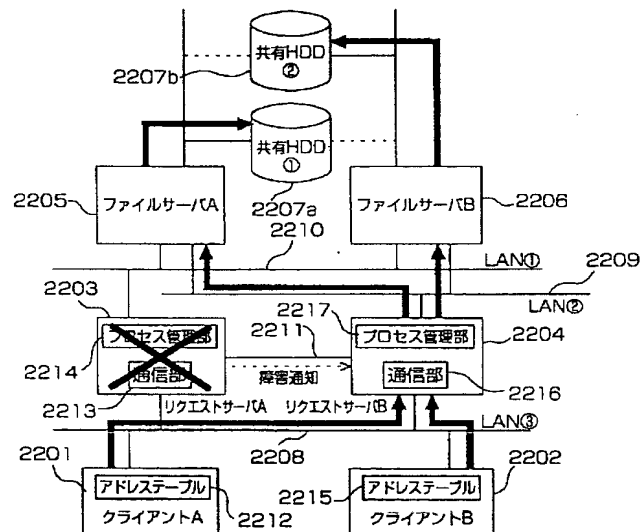
【図19】

【図 20】



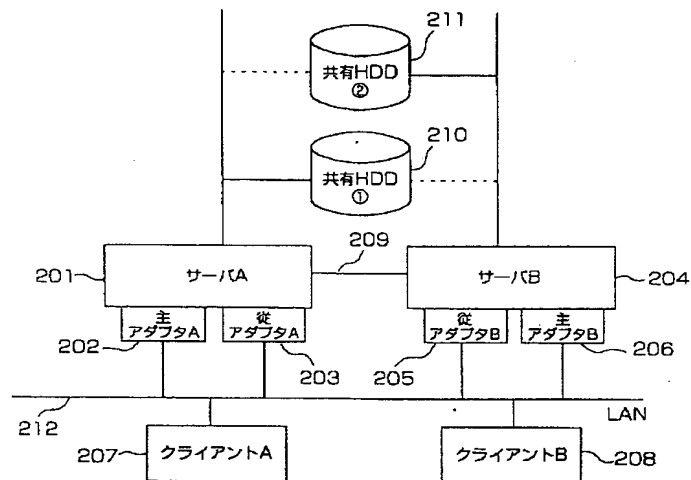
【図 20】

【図 21】



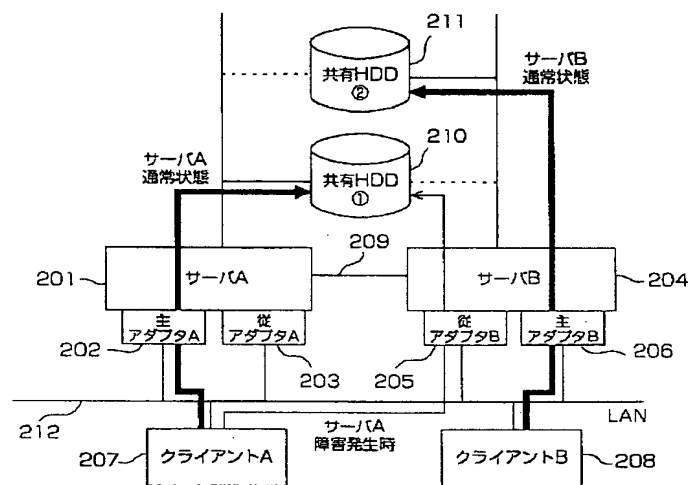
【図 21】

【図22】



【図22】

【図24】



【図24】

フロントページの続き

(72)発明者 宇賀神 敦
神奈川県海老名市下今泉810番地 株式会社
日立製作所オフィスシステム事業部内

(72)発明者 後藤 法宏
神奈川県横浜市戸塚区戸塚町5030番地 株
式会社日立製作所ソフトウェア開発本部内